

Research on the Application of Convolution Neural Networks in Handwritten Character Recognition and Face Recognition

Guodong Zhang

University of Science & Technology Beijing, USTB, Beijing, China

chinazhangguodong@163.com

Abstract: To study the convolution neural network and promote the research of image processing, the development and research achievements of artificial neural network and convolution neural network at home and abroad were summarized, and the Research on artificial neural networks, especially convolution neural networks in handwritten character recognition and face recognition. Based on the canonical convolution neural network model, it was applied to handwritten numeral recognition and face recognition tasks. The experiment proved that the convolution neural network could be applied to the problem of handwritten digital recognition and face recognition without too much adjustment and modification. It has achieved good recognition effect.

Keywords: convolutional neural network; artificial neural network; handwritten digital recognition; face recognition.

1. INTRODUCTION

Convolution neural network is a class of multilayer neural networks specially designed for processing two-dimensional data. The design of convolution neural networks is inspired by the early time-delay neural network. Time delay networks are used to deal with speech and time series signals, and the sharing weight on the time dimension is used to reduce the computational complexity in the learning process. Convolution neural network is considered to be the first truly successful deep learning method using multilayer hierarchical network. The convolution neural network can reduce the number of training parameters in the network by mining the spatial correlation in the data, and reach the efficiency of the network back-propagation algorithm, because the convolution neural network needs very little data preprocessing work, so it is also considered as a method of deep learning. In the convolution neural network, the small area of the image (also called "local perception area") is regarded as the underlying input data in the hierarchical structure. The information is propagated through the various layers in the network, and the filter is made up of the filter in each layer, so that some significant features of the measured data can be obtained. Because the local sensing area can obtain some basic features, such as the boundary and corner in the image, this method can provide a certain level of relative invariance to the displacement, stretching and rotation.

Image recognition refers to the technique of digital image processing, which uses artificial intelligence technology, especially the method of machine learning so that the computer can recognize the content of the image. Image recognition is a major field of pattern recognition, which involves handwritten character recognition, face recognition, object recognition and so on, and some of the relatively mature technologies have been used in commercial applications. At the same time, with the development of deep learning technology, the deep network also shows the characteristics of the universal image recognition system and has begun to be applied. Perez proposed a genetic algorithm to select the perceptual field model in the network, and also achieved good results in image recognition.

2. LITERATURE REVIEW

In the past decades, how to imitate the ability of human brain to express information efficiently has become a core issue in the field of artificial intelligence. Human beings receive a large amount of perceived information data at every moment, but they can accurately capture the key factors in these data and save them for future use. Through the study of neuroanatomy, neurophysiology, and the electrophysiological processes of neurons, a breakthrough has been made in the last century. People have become more and more aware of the structure and basic unit of the human brain, and researchers have begun to try to mimic the structure of the human brain. It works to construct algorithms that can recognize and memorize functions similar to human brain. The learning system of the brain is an extremely complex network system composed of interconnected neurons. In 1943, McCulloch and Pitts proposed the first individual working neuron model, using the simplified signal propagation mechanism to mimic some basic functions of the human brain neurons, thus laying a foundation for the basis of the development of early neural computing [1]. In 1958, on the basis of the MP model, Rosenblatt added the learning mechanism, and proposed a perceptron model to solve some problems in the field of character recognition. The first time the artificial neural network theory was applied to the actual problem, and it was proved that the network composed of two layers of perceptron could carry on the linear classification operation of the input. A very important research direction [2] with hidden layer perceptron units is also proposed. But after that, because of the rapid development of computer hardware and other fields, and the failure of artificial neural network to deal with nonlinear classification, the field has been in a long period of calm period.

3. METHODOLOGY

3.1 Research methods of artificial and convolution neural network

The BPANN network is a directed acyclic network structure formed by layered interconnection of some perceptron units, as shown in Figure 1. Each circle in the graph represents a perceptron unit. The arrow with the direction represents the connection between the sensor units, and the perceptron unit at the same level in the network constitutes a network layer. Generally speaking, in addition to one input layer and one output layer, the network also contains one or more hidden cell layer [3]. The input data received by the input layer propagate layer by layer after each hidden layer and reach the output layer.

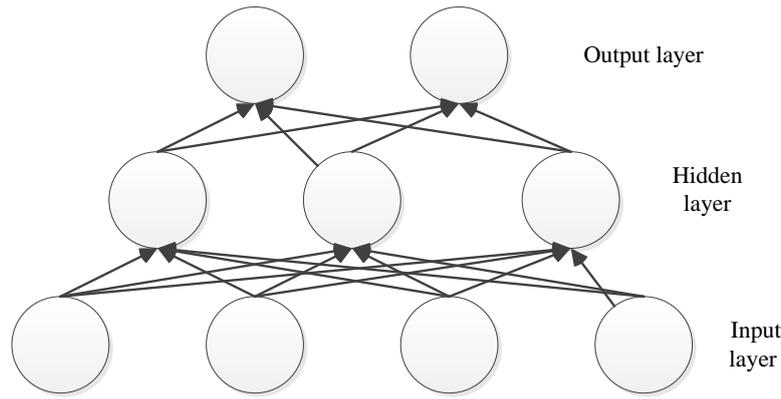


Figure.1 Simple multi-layer feed-forward network structure

The feed-forward process and the reverse propagation process alternate, until the output results of the network reach the predefined condition, usually the error is less than a certain set value or the number of iterations to a certain set value [4]. For a multi class classification problem with C classes and N training samples, the error function is given by the following formula:

$$E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (1)$$

Since the error of the entire data set is the sum of all single data errors, we need to consider the back-propagation algorithm [5] relative to a single example, for example, the n input:

$$E^n = \frac{1}{2} \sum_{k=1}^c (t_k^n - y_k^n)^2 = \frac{1}{2} \|t^n - y^n\|_2^2 \quad (2)$$

3.2 Application of convolution neural network in handwritten numeral recognition and face recognition

Application on handwriting recognition:

In this handwritten numeral recognition task, the paper adopts a network structure similar to LeNet-5. The input data is a matrix consisting of the pixels of 32*32. The first feature layer (C1) contains 6 feature graphs, and the input image is convoluted with the window of the 5*5, and the size of each feature graph is 28*28. Then the first down sampling level (S2) is used to perform the down sampling operation for C1, and the same 6 feature maps are obtained, but the size is reduced to 14*14. The C3 layer is a volume layer. The size of the convolution kernel is also 5*5. The connection between the C3 layer and the S2 layer is shown in Table 1. S4 is a descending [6] on the basis of C3. The S4 layer is convoluted by the C5 layer. The convolution kernel in each C5 is convoluted on all the feature maps of S4.

Table 1 LeNet-5middle layer and layer connection mode

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	○				○	○	○			○	○	○	○		○	○
1	○	○				○	○	○			○	○	○	○		○
2	○	○	○				○	○	○			○		○	○	○
3		○	○	○			○	○	○	○			○		○	○
4			○	○	○			○	○	○	○		○	○		○
5				○	○	○			○	○	○	○		○	○	○

The experimental results show that CNN1-3 is also based on CNN1-1 and increases the number of filters in each layer of the network. The experimental results also show that in the case of 400 training samples, the CNN1-3 cannot converge well in the training process, and the classification recognition rate is worse than the other network models.

Application of convolution neural network to face recognition

There are 40 categories in the ORL face recognition task, so the LeNet-5 network model is no longer applicable and needs to be adjusted appropriately. The network model used in this experiment is based on CNN1-1, CNN1-2 and CNN1-3 in 3.3. The number of units in the output layer is increased from 10 to 40, corresponding to the number of classified categories, the same network structure is also connected with the last layer, and the rest of the convolution neural network keeps the same [7]. In order to distinguish between the convolution neural network model used before, the convolution neural network models used in the ORL face data set are named CNN2-1, CNN2-2 and CNN2-3 respectively.

Table 2 compares the recognition rates of some commonly used face recognition methods on ORL datasets. It should be pointed out that all these methods employ a large number of image pre-processing methods to extract the features in the image beforehand. The experiment uses the direct original image input, and the recognition performance is good, and the convolution neural network as a kind of image recognition method has a wide range of features [8].

Table 2 Some common face recognition methods are classified

Classifier	Recognition rate (%)
Top down hidden markov model	13
Characteristics of the face	10.5
SOM+ convolutional neural network	3.8
Image sampling extraction algorithm + convolutional neural network	5.75
Multilayer perceptron network	39.6

Experimental comparison shows that CNN2-3 cannot converge to network parameters after training, and the classification recognition rate is much lower than that of other network models.

4. RESULTS AND DISCUSSION

In the paper, the following conclusions are drawn from the research of handwritten digit recognition and face recognition based on convolution neural network:

First, CNN1-1 is a simplified LeNet-5 network structure. Experimental results show that the simplified network can also achieve a good classification rate. On the basis of CNN1-1, a better classification recognition rate can be obtained by appropriately reducing the number of filters in each layer of the network, and the learning speed is improved to a certain extent, and the number of iterations needed in the network training process is less. CNN1-3 is also based on CNN1-1 and increases the number of filters in each layer of the network. It also shows that in the case of 400 training samples, the CNN1-3 cannot converge well in the training process, and the classification recognition rate is worse than the other network models.

Second, Table 3 compares the data in the above sections. The experimental results show that the CNN2-1 model can get a good classification and recognition rate on the ORL face dataset. Compared

with CNN2-1, CNN2-2 can get better classification and recognition rate with less learning time. After training, the network parameters of CNN2-3 cannot be converged, and the classification recognition rate is much lower than that of other network models.

Table 3 Several learning performance comparisons of three network structures on ORL data sets

	Training MCR (%)	Test MCR (%)	Number of convergent iterations
CNN2-1	2	9	18
CNN2-2	2	10.5	15
CNN2-3	14	23	Can't convergence

5. CONCLUSION

The deep network represented by convolution neural network has brought new research focus to artificial neural network for many years. The structure of the convolution neural network has the structure of local perception and hierarchical structure, which can effectively extract and classify the characteristics of the traditional recognition technology in two stages together and use the back-propagation algorithm to achieve the global training process. General image recognition is a new recognition task that has recently developed with the development of computer hardware and artificial intelligence technology. It is expected to be able to find some more general problems which are not based on specific ones. The research results of deep network in this class of problems are quite significant, and there is no need for too many different recognition tasks that can adjust the network structure and parameters. Handwritten digital recognition and face recognition are two distinct features of recognition. The number of handwritten numerals is less and the number of samples is sufficient, while the face data is just the opposite. These two recognition tasks are usually used as the main subjects of general image recognition problems.

REFERENCES

- [1] Bas E. The training of multiplicative neuron model based artificial neural networks with differential evolution algorithm for forecasting. *Journal of Artificial Intelligence and Soft Computing Research*, 2016, 6(1), pp. 5-11.
- [2] Feng X, Feuvrie B, Descamps A S, et al. Digital predistortion technique based on non-uniform MP model and interpolated LUT for linearising PAs with memory effects. *Electronics Letters*, 2014, 50(24), pp. 1882-1884.
- [3] Yilmaz M, Turgut B, Gullu M, et al. Application of artificial neural networks to height transformation. *Tehnički vjesnik*, 2017, 24(2), pp. 443-448.
- [4] Mirjalili S, Mirjalili S M, Lewis A. Let a biogeography-based optimizer train your multi-layer perceptron. *Information Sciences*, 2014, 269, pp. 188-209.
- [5] Wang L, Zeng Y, Chen T. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications*, 2015, 42(2), pp. 855-863.
- [6] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, 15(1), pp. 1929-1958.
- [7] Li G, Zhou B, Su Y. Face Recognition Algorithm Using Two Dimensional Locality Preserving Projection in Discrete Wavelet Domain. *Open Automation and Control Systems Journal*, 2015, 7, pp. 1721-1728.
- [8] Xu Y, Zhang Z, Lu G, et al. Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification *Pattern Recognition*, 2016, 54, pp. 68-82.